Convergence of a Smoothing Continuation Method for Mathematical Programs with Complementarity Constraints

Masao Fukushima* Jong-Shi Pang†

March 10, 1999

Abstract

We study the behavior of a sequence generated by a smoothing continuation method for mathematical programs with equilibrium constraints (MPEC). In particular, we show that under the linear independence constraint qualification and an additional condition called the asymptotic weak nondegeneracy, the limit of KKT points satisfying the second-order necessary conditions for the perturbed problems is a B-stationary point of the original MPEC. The notable fact is that this result does not rely on the strict complementarity assumption at the limit point.

Keywords: Mathematical programs with equilibrium constraints, stationary point, constraint qualification, smoothing continuation method.

AMS subject classification: 90C30, 90C33

1 Introduction

We consider the following mathematical programming problem with equilibrium constraints (MPEC):

$$\begin{array}{lll} \text{P:} & \text{minimize} & f(z) \\ & \text{subject to} & G(z) \geq 0, \quad H(z) \geq 0, \quad G(z)^T H(z) = 0, \\ & & g(z) \leq 0, \quad h(z) = 0, \end{array}$$

where $f : \Re^n \to \Re$, $G : \Re^n \to \Re^m$, $H : \Re^n \to \Re^m$, $g : \Re^n \to \Re^p$ and $h : \Re^n \to \Re^q$ are twice continuously differentiable functions. Although this formulation looks somewhat different from the conventional MPECs that involve the first-level

*Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan, Email: fuku@kuamp.kyoto-u.ac.jp

†Department of Mathematical Sciences, Whiting School of Engineering, The Johns Hopkins University, Baltimore, Maryland 21218-2692, U.S.A., Email: jsp@vicp1.mts.jhu.edu

and the second-level variables, the latter problems can be transformed into the form of P in a straightforward manner.

The major cause of difficulty in dealing with MPECs lies in the unusual nature of the constraints. For example, it is known [1, 12] that a standard constraint qualifications in nonlinear programming such as Mangasarian-Fromovitz constraint qualification (MFCQ) never holds for problem P. Nevertheless, serious attention has recently been paid to clarifying stationarity concepts for various types of MPECs under novel constraint qualifications pertaining to the peculiar structure of equilibrium constraints [7, 8, 11, 12, 13]. On the other hand, the development of efficient and robust algorithms for MPECs seems somewhat behind the theoretical study, although attempts have been made quite intensively in recent years [2, 4, 5, 7, 9, 10]. The algorithms proposed so far are only guaranteed to converge to a point satisfying some weak stationarity conditions such as Clarke stationarity condition, or require the nondegeneracy (strict complementarity) assumption at a limit point so as to ensure a stronger stationarity concept such as B-stationarity. The main theme of this paper is to establish convergence to a B-stationary point for a smoothing continuation method without the nondegeneracy assumption. The basic assumptions we adopt here are the linear independence constraint qualification (LICQ) and an additional condition called the asymptotic weak nondegeneracy at a limit of the generated sequence. Based on the fact that, under LICQ, a KKT point for a relaxed problem associated with the given MPEC affords a B-stationary point of the MPEC, we will show that the generated sequence converges to a KKT point for the relaxed problem.

Notations: The component functions of G, H, g and h are denoted by $G_i, i = 1, \ldots, m$, $H_i, i = 1, \ldots, m$, $g_l, l = 1, \ldots, p$, and $h_k, k = 1, \ldots, q$, respectively. For a vector $v \in \Re^m$ and index set $\alpha \in \{1, \ldots, m\}$, v_α denotes the subvector of v with components $v_i, i \in \alpha$. Similar notations are used for vector valued functions as well. The index sets $\{1, \ldots, p\}$ and $\{1, \ldots, q\}$ are denoted by $\mathcal{I}$ and $\mathcal{J}$, respectively.

2 Preliminaries

Let $\bar{z}$ be a given feasible point of problem P. We define the index sets

$$\begin{array}{rcl} \bar{\alpha} & \equiv & \{\, i \,|\, G_i(\bar{z}) = 0 < H_i(\bar{z}) \,\} \\ \bar{\beta} & \equiv & \{\, i \,|\, G_i(\bar{z}) = 0 = H_i(\bar{z}) \,\} \\ \bar{\gamma} & \equiv & \{\, i \,|\, G_i(\bar{z}) > 0 = H_i(\bar{z}) \,\} \\ \bar{\mathcal{I}} & \equiv & \{\, l \,|\, g_l(\bar{z}) = 0 \,\}. \end{array} \tag{1}$$

In particular, we say that $\bar{z}$ is *nondegenerate* [7] if $\bar{\beta} = \emptyset$, i.e., $\bar{z}$ satisfies the *strict complementarity* condition $G_i(\bar{z}) + H_i(\bar{z}) > 0$ for all $i = 1, \ldots, m$. Associated

with the index sets $\bar{\alpha}, \bar{\beta}$ and $\bar{\gamma}$, we define the *relaxed problem* of P at $\bar{z}$:

$$\begin{array}{lll} \mathrm{R}(\bar{z}): & \text{minimize} & f(z) \\ & \text{subject to} & G_i(z) = 0, \quad i \in \bar{\alpha} \\ & & G_i(z) \geq 0, \quad i \in \bar{\beta} \\ & & H_i(z) \geq 0, \quad i \in \bar{\beta} \\ & & H_i(z) = 0, \quad i \in \bar{\gamma} \\ & & g_l(z) \leq 0, \quad l \in \mathcal{I} \\ & & h_k(z) = 0, \quad k \in \mathcal{J}. \end{array}$$

We say that the *linear independence constraint qualification (LICQ)* holds at $\bar{z}$ if the vectors

$$\begin{array}{c} \{ \nabla G_i(\bar{z}) \,|\, i \in \bar{\alpha} \cup \bar{\beta} \} \cup \{ \nabla H_i(\bar{z}) \,|\, i \in \bar{\beta} \cup \bar{\gamma} \} \\ \cup \{ \nabla g_l(\bar{z}) \,|\, l \in \bar{\mathcal{I}} \} \cup \{ \nabla h_k(\bar{z}) \,|\, k \in \mathcal{J} \} \end{array}$$

are linearly independent.
The *Karush-Kuhn-Tucker (KKT) conditions* for the relaxed problem $\mathrm{R}(\bar{z})$ are represented as

$$\begin{array}{c} \nabla f(z) - \sum_{i \in \bar{\alpha}} v_i \nabla G_i(z) - \sum_{i \in \bar{\beta}} v_i \nabla G_i(z) - \sum_{i \in \bar{\beta}} w_i \nabla H_i(z) - \sum_{i \in \bar{\gamma}} w_i \nabla H_i(z) \\ + \sum_{l \in \mathcal{I}} \lambda_l \nabla g_l(z) + \sum_{k \in \mathcal{J}} \mu_k \nabla h_k(z) \;=\; 0, \\ v_{\bar{\beta}} \geq 0, \quad G_{\bar{\beta}}(z) \geq 0, \quad v_{\bar{\beta}}^T G_{\bar{\beta}}(z) = 0, \\ w_{\bar{\beta}} \geq 0, \quad H_{\bar{\beta}}(z) \geq 0, \quad w_{\bar{\beta}}^T H_{\bar{\beta}}(z) = 0, \\ \lambda \geq 0, \quad g(z) \leq 0, \quad \lambda^T g(z) = 0, \\ G_{\bar{\alpha}}(z) = 0, \quad H_{\bar{\gamma}}(z) = 0, \quad h(z) = 0. \end{array} \tag{2}$$

Let S be an arbitrary subset of $\Re^n$. We call $\bar{z} \in S$ a *B-stationary point* of the minimization problem

$$\begin{array}{ll} \text{minimize} & f(z) \\ \text{subject to} & z \in S \end{array}$$

if it satisfies

$$\nabla f(\bar{z})^T d \geq 0, \quad \forall\, d \in \mathcal{T}(\bar{z}; S),$$

where $\mathcal{T}(\bar{z}; S)$ is the tangent cone of S at $\bar{z}$ that consists of all vectors d for which there exist a sequence $\{z^\nu\} \subset S$ converging to $\bar{z}$ and a sequence $\{\tau_\nu\}$ of positive numbers converging to zero such that $d = \lim_{\nu \to \infty} (z^\nu - \bar{z})/\tau_\nu$. Clearly any local optimal solution is a B-stationary point.

Under some constraint qualification, any B-stationary point of the relaxed problem R($\bar{z}$) satisfies the KKT conditions (2) along with some Lagrange multipliers $v_{\bar{\alpha}\cup\bar{\beta}}$, $w_{\bar{\beta}\cup\bar{\gamma}}$, λ and μ. Furthermore, if the LICQ holds at $\bar{z}$, then a B-stationary point of R($\bar{z}$) is also a B-stationary point for problem P [7, Proposition 4.3.7].

Proposition 2.1 *If a feasible point $\bar{z}$ of problem P satisfies the KKT conditions (2) for the relaxed problem R($\bar{z}$), and if the LICQ holds at $\bar{z}$, then $\bar{z}$ is a B-stationary point of problem P.*

3 Main Results

We consider a smoothing continuation method for solving problem P. This method uses the perturbed Fischer-Burmeister function $\phi : \Re^2 \times \Re_+ \to \Re$ defined by

$$\phi(a,b,\varepsilon) = a + b - \sqrt{a^2 + b^2 + \varepsilon}. \tag{3}$$

The function ϕ was first introduced by Kanzow [6] for linear complementarity problems, and used in particular to develop an SQP method for solving MPEC in [4]. When $\varepsilon = 0$, the function ϕ reduces to the Fischer-Burmeister function and satisfies

$$\phi(a,b,0) = 0 \quad \Longleftrightarrow \quad a = 0,\ b = 0,\ ab = 0. \tag{4}$$

When $\varepsilon > 0$, this function satisfies

$$\phi(a,b,\varepsilon) = 0 \quad \Longleftrightarrow \quad a > 0,\ b > 0,\ ab = \varepsilon/2 \tag{5}$$

and is smooth with respect to (a,b). More specifically, its first and second derivatives are given by

$$\begin{aligned}
\frac{\partial}{\partial a}\phi(a,b,\varepsilon) &= 1 - \frac{a}{\sqrt{a^2 + b^2 + \varepsilon}}, \\
\frac{\partial}{\partial b}\phi(a,b,\varepsilon) &= 1 - \frac{b}{\sqrt{a^2 + b^2 + \varepsilon}}, \\
\frac{\partial^2}{\partial a^2}\phi(a,b,\varepsilon) &= -\frac{b^2 + \varepsilon}{(a^2 + b^2 + \varepsilon)^{3/2}}, \\
\frac{\partial^2}{\partial b^2}\phi(a,b,\varepsilon) &= -\frac{a^2 + \varepsilon}{(a^2 + b^2 + \varepsilon)^{3/2}}, \\
\frac{\partial^2}{\partial a \partial b}\phi(a,b,\varepsilon) &= \frac{ab}{(a^2 + b^2 + \varepsilon)^{3/2}}.
\end{aligned} \tag{6}$$

With the perturbed Fischer-Burmeister function ϕ, we define the function $\Phi^\varepsilon : \Re^n \to \Re^m$ by

$$\Phi^\varepsilon(z) = \begin{pmatrix} \phi(G_1(z), H_1(z), \varepsilon) \\ \vdots \\ \phi(G_m(z), H_m(z), \varepsilon) \end{pmatrix}. \tag{7}$$

For $\varepsilon > 0$, the function Φ^ε is differentiable everywhere and the gradient of its component function Φ_i^ε at z is written as

$$\begin{aligned}\nabla \Phi_i^\varepsilon(z) &= \left(1 - \frac{G_i(z)}{\sqrt{G_i(z)^2 + H_i(z)^2 + \varepsilon}}\right) \nabla G_i(z) \\ &\quad + \left(1 - \frac{H_i(z)}{\sqrt{G_i(z)^2 + H_i(z)^2 + \varepsilon}}\right) \nabla H_i(z). \end{aligned} \tag{8}$$

In particular, we have the following lemma that will be useful in subsequent discussions.

Lemma 3.1 *Let z satisfy $\Phi^\varepsilon(z) = 0$ with $\varepsilon > 0$. Then we have*

$$G_i(z) > 0, \quad H_i(z) > 0, \quad G_i(z) H_i(z) = \varepsilon/2$$

for all $i = 1, \ldots, m$. Moreover we have

$$\begin{aligned}
\frac{\partial}{\partial a}\phi(G_i(z), H_i(z), \varepsilon) &= \frac{H_i(z)}{G_i(z) + H_i(z)}, \\
\frac{\partial}{\partial b}\phi(G_i(z), H_i(z), \varepsilon) &= \frac{G_i(z)}{G_i(z) + H_i(z)}, \\
\frac{\partial^2}{\partial a^2}\phi(G_i(z), H_i(z), \varepsilon) &= -\frac{H_i(z)(H_i(z) + 2G_i(z))}{(G_i(z) + H_i(z))^3}, \\
\frac{\partial^2}{\partial b^2}\phi(G_i(z), H_i(z), \varepsilon) &= -\frac{G_i(z)(G_i(z) + 2H_i(z))}{(G_i(z) + H_i(z))^3}, \\
\frac{\partial^2}{\partial a \partial b}\phi(G_i(z), H_i(z), \varepsilon) &= \frac{G_i(z) H_i(z)}{(G_i(z) + H_i(z))^3}
\end{aligned}$$

for all $i = 1, \ldots, m$.

Proof. The first part is immediate from (5). Then we have

$$\begin{aligned}
\sqrt{G_i(z)^2 + H_i(z)^2 + \varepsilon} &= \sqrt{G_i(z)^2 + H_i(z)^2 + 2G_i(z)H_i(z)} \\
&= |G_i(z) + H_i(z)| \\
&= G_i(z) + H_i(z).
\end{aligned}$$

This together with (6) yields the latter part of the lemma. □

By Lemma 3.1, the gradient of Φ_i^ε at a point z such that $\Phi^\varepsilon(z) = 0$ is particularly given by

$$\nabla \Phi_i^\varepsilon(z) = \frac{H_i(z)}{G_i(z) + H_i(z)} \nabla G_i(z) + \frac{G_i(z)}{G_i(z) + H_i(z)} \nabla H_i(z). \tag{9}$$

For $\varepsilon = 0$, the function Φ^0 is not differentiable in general. Specifically, when $G_i(\bar{z}) = H_i(\bar{z}) = 0$ for some $\bar{z}$ and i, i.e., $i \in \bar{\beta}$, the function Φ_i^0 is not differentiable at $\bar{z}$ and the Clarke generalized gradient

$$\partial \Phi_i^0(\bar{z}) \equiv \{r \in \Re^n \,|\, r = \lim_{k \to \infty} \nabla \Phi_i^0(z^k) \text{ with } z^k \to \bar{z} \text{ and } \nabla \Phi_i^0(z^k) \text{ exist}\}$$

is contained in the set

$$\partial_C \Phi_i^0(\bar{z}) \equiv \{r \in \Re^n \,|\, r = \bar{\xi}_i \nabla G_i(\bar{z}) + \bar{\eta}_i \nabla H_i(\bar{z}) \text{ for some } (\bar{\xi}_i, \bar{\eta}_i) \in B\}, \quad (10)$$

where $B \equiv \{(\bar{\xi}_i, \bar{\eta}_i) \in \Re^2 \,|\, (1 - \bar{\xi}_i)^2 + (1 - \bar{\eta}_i)^2 \leq 1\}$.
Now consider the following perturbed problem for P with $\varepsilon > 0$:

$$\begin{array}{lll} \mathrm{P}^\varepsilon\colon & \text{minimize} & f(z) \\ & \text{subject to} & \Phi^\varepsilon(z) = 0, \\ & & g(z) \leq 0, \quad h(z) = 0. \end{array}$$

We have the following lemma.

Lemma 3.2 *For each $\varepsilon > 0$, let z^ε be a feasible point of P^ε. Suppose that $z^\varepsilon \to \bar{z}$ as $\varepsilon \to 0$. Then $\bar{z}$ is a feasible point of P. Moreover, if the LICQ holds at $\bar{z}$, then the gradients*

$$\{\, \nabla \Phi_i^\varepsilon(z^\varepsilon) \,|\, i = 1, \ldots, m \,\} \;\cup\; \{\, \nabla g_l(z^\varepsilon) \,|\, l \in \mathcal{I}^\varepsilon \,\} \;\cup\; \{\, \nabla h_k(z^\varepsilon) \,|\, k \in \mathcal{J} \,\}$$

are linearly independent for all $\varepsilon > 0$ small enough, where $\mathcal{I}^\varepsilon \equiv \{\, l \,|\, g_l(z^\varepsilon) = 0 \,\}$.

Proof. By (4), (5) and the continuity of ϕ, it is easy to see that $\bar{z}$ is feasible to P. Let index sets $\bar{\alpha}, \bar{\beta}, \bar{\gamma}, \bar{\mathcal{I}}$ be defined by (1). Then it follows from (8) that as $\varepsilon \to 0$

$$\begin{array}{rcll} \|\nabla \Phi_i^\varepsilon(z^\varepsilon) - \nabla G_i(\bar{z})\| & \to & 0 & \text{for } i \in \bar{\alpha}, \\ \|\nabla \Phi_i^\varepsilon(z^\varepsilon) - \nabla H_i(\bar{z})\| & \to & 0 & \text{for } i \in \bar{\gamma}, \\ \mathrm{dist}[\nabla \Phi_i^\varepsilon(z^\varepsilon), \partial_C \Phi_i^0(\bar{z})] & \to & 0 & \text{for } i \in \bar{\beta}, \end{array}$$

where $\mathrm{dist}[y, S] \equiv \min\{\|y - y'\| \,|\, y' \in S\}$ for $y \in \Re^n$ and $S \subset \Re^n$. Note that the set $\partial_C \Phi_i^0(\bar{z})$ contains the Clarke generalized gradient $\partial \Phi_i^0(\bar{z})$ of Φ_i^0 at $\bar{z}$. Moreover, by continuity, it is clear that

$$\begin{array}{rcll} \|\nabla g_l(z^\varepsilon) - \nabla g_l(\bar{z})\| & \to & 0 & \text{for } l \in \mathcal{I}, \\ \|\nabla h_k(z^\varepsilon) - \nabla h_k(\bar{z})\| & \to & 0 & \text{for } k \in \mathcal{J}. \end{array}$$

Observe that $\mathcal{I}^\varepsilon \subseteq \bar{\mathcal{I}}$ for all $\varepsilon > 0$ sufficiently small. Then it is not difficult to deduce that the desired result holds under the LICQ at $\bar{z}$. □

Now let z^ε be a local optimal solution of problem P^ε. Then, since the perturbed problem P^ε is an ordinary nonlinear programming problem, under some constraint qualification, there exist Lagrange multipliers $u^\varepsilon \in \Re^m, \lambda^\varepsilon \in \Re^p, \mu^\varepsilon \in \Re^q$

such that the vector $(z,u,\lambda,\mu) = (z^\varepsilon,u^\varepsilon,\lambda^\varepsilon,\mu^\varepsilon)$ satisfies the following KKT conditions for problem P^ε:

$$\begin{aligned}
&\nabla f(z) - \sum_{i=1}^{m} u_i \nabla \Phi_i^\varepsilon(z) + \sum_{l\in\mathcal{I}} \lambda_l \nabla g_l(z) + \sum_{k\in\mathcal{J}} \mu_k \nabla h_k(z) = 0,\\
&\Phi^\varepsilon(z) = 0,\\
&\lambda \geq 0, \quad g(z) \leq 0, \quad \lambda^T g(z) = 0,\\
&h(z) = 0.
\end{aligned} \tag{11}$$

Moreover $(z,u,\lambda,\mu) = (z^\varepsilon,u^\varepsilon,\lambda^\varepsilon,\mu^\varepsilon)$ satisfies the inequalities

$$d^T \nabla_z^2 \mathcal{L}^\varepsilon(z,u,\lambda,\mu) d \geq 0 \quad \text{for all} \quad d \in \mathcal{T}^\varepsilon(z), \tag{12}$$

where

$$\mathcal{L}^\varepsilon(z,u,\lambda,\mu) \equiv f(z) - \sum_{i=1}^{m} u_i \Phi_i^\varepsilon(z) + \sum_{l\in\mathcal{I}} \lambda_l g_l(z) + \sum_{k\in\mathcal{J}} \mu_k h_k(z)$$

and

$$\mathcal{T}^\varepsilon(z) \equiv \{\, d \in \Re^n \,|\, \nabla\Phi^\varepsilon(z)^T d = 0,\ \nabla g_{\mathcal{I}_\varepsilon}(z)^T d = 0,\ \nabla h(z)^T d = 0 \,\}$$

with $\mathcal{I}_\varepsilon \equiv \{\, l \,|\, g_l(z^\varepsilon) = 0 \,\}$.

To establish our main result, we introduce the concept of asymptotic weak nondegeneracy. Suppose that $z^\varepsilon \to \bar{z}$ as $\varepsilon \to 0$. It then follows from (9) that, for each $i \in \bar{\beta}$, any accumulation point $\bar{r}$ of $\{\nabla\Phi_i^\varepsilon(z^\varepsilon)\}$ belongs to the set $\partial_C \Phi_i^0(\bar{z})$ defined by (10) and hence is represented as

$$\bar{r} = \bar{\xi}_i \nabla G_i(\bar{z}) + \bar{\eta}_i \nabla H_i(\bar{z})$$

for some $(\bar{\xi}_i, \bar{\eta}_i)$ such that $(1-\bar{\xi}_i)^2 + (1-\bar{\eta}_i)^2 \leq 1$. We say that $\{z^\varepsilon\}$ is *asymptotically weakly nondegenerate* if neither $\bar{\xi}_i$ nor $\bar{\eta}_i$ vanishes for any accumulation point $\bar{r}$. Roughly speaking, in view of (9), this condition means that $G_i(z^\varepsilon)$ and $H_i(z^\varepsilon)$ approach zero in the same order of magnitude. Obviously $\{z^\varepsilon\}$ is asymptotically weakly nondegenerate when $\bar{z}$ satisfies the nondegeneracy condition $\bar{\beta} = \emptyset$.
Now we state the main theorem.

Theorem 3.1 *Let $(z^\varepsilon,u^\varepsilon,\lambda^\varepsilon,\mu^\varepsilon)$ satisfy the second-order necessary conditions (11) and (12) for problem P^ε. Suppose that $(z^\varepsilon,u^\varepsilon,\lambda^\varepsilon,\mu^\varepsilon) \to (\bar{z},\bar{u},\bar{\lambda},\bar{\mu})$ as $\varepsilon \to 0$. If the LICQ holds at $\bar{z}$ and $\{z^\varepsilon\}$ is asymptotically weakly nondegenerate, then $\bar{z}$ satisfies the KKT conditions (2) for the relaxed problem R($\bar{z}$); moreover $\bar{z}$ is a B-stationary point for P.*

Proof. For each $i \in \bar{\beta}$, let $\bar{r}$ be an arbitrary accumulation point of $\{\nabla \Phi_i^\varepsilon(z^\varepsilon)\}$. Then, by the asymptotic weak nondegeneracy assumption, we have

$$\bar{r} = \bar{\xi}_i \nabla G_i(\bar{z}) + \bar{\eta}_i \nabla H_i(\bar{z})$$

with $(1-\bar{\xi}_i)^2 + (1-\bar{\eta}_i)^2 \leq 1$ and $\bar{\xi}_i > 0$ and $\bar{\eta}_i > 0$. Since

$$\begin{aligned} \nabla \Phi_i^\varepsilon(z^\varepsilon) &\rightarrow \nabla G_i(\bar{z}) \quad \text{for } i \in \bar{\alpha}, \\ \nabla \Phi_i^\varepsilon(z^\varepsilon) &\rightarrow \nabla H_i(\bar{z}) \quad \text{for } i \in \bar{\gamma}, \end{aligned}$$

it follows from the first equation in the KKT conditions (11) for problem P^ε that

$$\begin{aligned} \nabla f(\bar{z}) - \sum_{i \in \bar{\alpha}} \bar{u}_i \nabla G_i(\bar{z}) - \sum_{i \in \bar{\beta}} \bar{u}_i \bar{\xi}_i \nabla G_i(\bar{z}) - \sum_{i \in \bar{\beta}} \bar{u}_i \bar{\eta}_i \nabla H_i(\bar{z}) - \sum_{i \in \bar{\gamma}} \bar{u}_i \nabla H_i(\bar{z}) \\ + \sum_{l \in \mathcal{I}} \bar{\lambda}_l \nabla g_l(\bar{z}) + \sum_{k \in \mathcal{J}} \bar{\mu}_k \nabla h_k(\bar{z}) \;=\; 0. \end{aligned} \tag{13}$$

It is easy to see that

$$\begin{aligned} \bar{\lambda}_l &\geq 0 \quad \text{for } l \in \bar{\mathcal{I}}, \\ \bar{\lambda}_l &= 0 \quad \text{for } l \notin \bar{\mathcal{I}}. \end{aligned}$$

Hence it suffices to show that

$$\begin{aligned} \bar{v}_i &\equiv \bar{u}_i \bar{\xi}_i \;\geq\; 0 \quad \text{for } i \in \bar{\beta}, \\ \bar{w}_i &\equiv \bar{u}_i \bar{\eta}_i \;\geq\; 0 \quad \text{for } i \in \bar{\beta}. \end{aligned} \tag{14}$$

We will derive a contradiction by assuming that $\bar{v}_j < 0$ for some $j \in \bar{\beta}$. We can argue the case where $\bar{w}_j < 0$ for some $j \in \bar{\beta}$ similarly. First note that, since $\bar{\xi}_i > 0$, $\bar{v}_j < 0$ implies $\bar{u}_j < 0$. Now let $\varepsilon > 0$ be sufficiently small and choose a vector $d^\varepsilon \in \Re^n$ such that

$$\begin{aligned} \nabla G_j(z^\varepsilon)^T d^\varepsilon &= \frac{\partial \phi_j^\varepsilon}{\partial b} \;\equiv\; \frac{\partial}{\partial b} \phi(G_j(z^\varepsilon), H_j(z^\varepsilon), \varepsilon), \\ \nabla H_j(z^\varepsilon)^T d^\varepsilon &= -\frac{\partial \phi_j^\varepsilon}{\partial a} \;\equiv\; -\frac{\partial}{\partial a} \phi(G_j(z^\varepsilon), H_j(z^\varepsilon), \varepsilon), \\ \nabla G_i(z^\varepsilon)^T d^\varepsilon &= 0 \qquad \text{for } i \in \bar{\alpha} \cup \bar{\beta} \setminus \{j\}, \\ \nabla H_i(z^\varepsilon)^T d^\varepsilon &= 0 \qquad \text{for } i \in \bar{\gamma} \cup \bar{\beta} \setminus \{j\}, \\ \nabla g_l(z^\varepsilon)^T d^\varepsilon &= 0 \qquad \text{for } l \in \mathcal{I}^\varepsilon, \\ \nabla h_k(z^\varepsilon)^T d^\varepsilon &= 0 \qquad \text{for } k \in \mathcal{J}. \end{aligned} \tag{15}$$

We can choose such d^ε because the LICQ at $\bar{z}$ ensures the linear independence of vectors

$$\begin{aligned} \{\, \nabla G_i(z^\varepsilon) \,|\, i \in \bar{\alpha} \cup \bar{\beta} \,\} \cup \{\, \nabla H_i(z^\varepsilon) \,|\, i \in \bar{\beta} \cup \bar{\gamma} \,\} \\ \cup \{\, \nabla g_l(z^\varepsilon) \,|\, l \in \mathcal{I}^\varepsilon \,\} \cup \{\, \nabla h_k(z^\varepsilon) \,|\, k \in \mathcal{J} \,\} \end{aligned}$$

for any $\varepsilon > 0$ small enough. (Note that $\mathcal{I}^\varepsilon \subseteq \bar{\mathcal{I}}$ for all ε small enough.) Then since

$$\nabla \Phi_j^\varepsilon(z^\varepsilon) = \frac{\partial \phi_j^\varepsilon}{\partial a} \nabla G_j(z^\varepsilon) + \frac{\partial \phi_j^\varepsilon}{\partial b} \nabla H_j(z^\varepsilon),$$

it follows from (15) that

$$\begin{aligned} \nabla \Phi_j^\varepsilon(z^\varepsilon)^T d^\varepsilon &= \frac{\partial \phi_j^\varepsilon}{\partial a} \nabla G_j(z^\varepsilon)^T d^\varepsilon + \frac{\partial \phi_j^\varepsilon}{\partial b} \nabla H_j(z^\varepsilon)^T d^\varepsilon \\ &= \frac{\partial \phi_j^\varepsilon}{\partial a} \frac{\partial \phi_j^\varepsilon}{\partial b} - \frac{\partial \phi_j^\varepsilon}{\partial b} \frac{\partial \phi_j^\varepsilon}{\partial a} \\ &= 0. \end{aligned}$$

Moreover we have

$$\begin{aligned} \nabla \Phi_i^\varepsilon(z^\varepsilon)^T d^\varepsilon &= 0 \qquad \text{for } i = 1, \ldots, m \ (i \neq j), \\ \nabla g_l(z^\varepsilon)^T d^\varepsilon &= 0 \qquad \text{for } l \in \mathcal{I}^\varepsilon, \\ \nabla h_k(z^\varepsilon)^T d^\varepsilon &= 0 \qquad \text{for } k \in \mathcal{J}. \end{aligned}$$

Hence it follows that

$$d^\varepsilon \in \mathcal{T}^\varepsilon(z^\varepsilon).$$

Now we consider the limit

$$\begin{aligned} &\lim_{\varepsilon \to 0} (d^\varepsilon)^T \nabla_z^2 \mathcal{L}(z^\varepsilon, u^\varepsilon, \lambda^\varepsilon, \mu^\varepsilon) d^\varepsilon \\ &= \lim_{\varepsilon \to 0} (d^\varepsilon)^T \left[\nabla^2 f(z^\varepsilon) - \sum_{i=1}^m u_i^\varepsilon \nabla^2 \Phi_i^\varepsilon(z^\varepsilon) \right. \\ &\qquad \left. + \sum_{l \in \mathcal{I}} \lambda_l^\varepsilon \nabla^2 g_l(z^\varepsilon) + \sum_{k \in \mathcal{J}} \mu_k^\varepsilon \nabla^2 h_k(z^\varepsilon) \right] d^\varepsilon. \end{aligned} \tag{16}$$

It follows from (15) that

$$\begin{aligned} &(d^\varepsilon)^T \nabla^2 \Phi_j^\varepsilon(z^\varepsilon) d^\varepsilon \\ &= (d^\varepsilon)^T \left(\frac{\partial \phi_j^\varepsilon}{\partial a} \nabla^2 G_j(z^\varepsilon) + \frac{\partial \phi_j^\varepsilon}{\partial b} \nabla^2 H_j(z^\varepsilon) \right. \\ &\quad + \frac{\partial^2 \phi_j^\varepsilon}{\partial a^2} \nabla G_j(z^\varepsilon) \nabla G_j(z^\varepsilon)^T + \frac{\partial^2 \phi_j^\varepsilon}{\partial a \partial b} \nabla G_j(z^\varepsilon) \nabla H_j(z^\varepsilon)^T \\ &\quad \left. + \frac{\partial^2 \phi_j^\varepsilon}{\partial b \partial a} \nabla H_j(z^\varepsilon) \nabla G_j(z^\varepsilon)^T + \frac{\partial^2 \phi_j^\varepsilon}{\partial b^2} \nabla H_j(z^\varepsilon) \nabla H_j(z^\varepsilon)^T \right) d^\varepsilon \end{aligned}$$

$$= \frac{\partial \phi_j^\varepsilon}{\partial a}(d^\varepsilon)^T \nabla^2 G_j(z^\varepsilon) d^\varepsilon + \frac{\partial \phi_j^\varepsilon}{\partial b}(d^\varepsilon)^T \nabla^2 H_j(z^\varepsilon) d^\varepsilon$$
$$+ \frac{\partial^2 \phi_j^\varepsilon}{\partial a^2}\left(\frac{\partial \phi_j^\varepsilon}{\partial b}\right)^2 - 2\frac{\partial^2 \phi_j^\varepsilon}{\partial a \partial b}\left(\frac{\partial \phi_j^\varepsilon}{\partial a}\right)\left(\frac{\partial \phi_j^\varepsilon}{\partial b}\right) + \frac{\partial^2 \phi_j^\varepsilon}{\partial b^2}\left(\frac{\partial \phi_j^\varepsilon}{\partial a}\right)^2 \quad (17)$$

and, for all $i \neq j$,

$$(d^\varepsilon)^T \nabla^2 \Phi_i^\varepsilon(z^\varepsilon) d^\varepsilon = \frac{\partial \phi_i^\varepsilon}{\partial a}(d^\varepsilon)^T \nabla^2 G_i(z^\varepsilon) d^\varepsilon + \frac{\partial \phi_i^\varepsilon}{\partial b}(d^\varepsilon)^T \nabla^2 H_i(z^\varepsilon) d^\varepsilon,$$

where

$$\frac{\partial^2 \phi_j^\varepsilon}{\partial a^2} \equiv \frac{\partial^2}{\partial a^2}\phi_j(G_j(z^\varepsilon), H_j(z^\varepsilon), \varepsilon),$$
$$\frac{\partial^2 \phi_j^\varepsilon}{\partial a \partial b} \equiv \frac{\partial^2}{\partial a \partial b}\phi_j(G_j(z^\varepsilon), H_j(z^\varepsilon), \varepsilon),$$
$$\frac{\partial^2 \phi_j^\varepsilon}{\partial b^2} \equiv \frac{\partial^2}{\partial b^2}\phi_j(G_j(z^\varepsilon), H_j(z^\varepsilon), \varepsilon).$$

Note that $\partial \phi_i^\varepsilon / \partial a$, $\partial \phi_i^\varepsilon / \partial b$, $\nabla^2 G_i(z^\varepsilon)$, $\nabla^2 H_i(z^\varepsilon)$ are all bounded as $\varepsilon \to 0$, since, by assumption, $(z^\varepsilon, u^\varepsilon, \lambda^\varepsilon, \mu^\varepsilon) \to (\bar{z}, \bar{u}, \bar{\lambda}, \bar{\mu})$. Also note that the LICQ at $\bar{z}$ and the boundedness of $\partial \phi_i^\varepsilon / \partial a$, $\partial \phi_i^\varepsilon / \partial b$ imply that d^ε is also bounded as $\varepsilon \to 0$. Therefore $(d^\varepsilon)^T \nabla^2 f(z^\varepsilon) d^\varepsilon$, $u_i^\varepsilon (d^\varepsilon)^T \nabla^2 \Phi_i^\varepsilon(z^\varepsilon) d^\varepsilon$ $(i \neq j)$, $\lambda_l^\varepsilon (d^\varepsilon)^T \nabla^2 g_l(z^\varepsilon) d^\varepsilon$ and $\mu_k^\varepsilon (d^\varepsilon)^T \nabla^2 h_k(z^\varepsilon) d^\varepsilon$ are bounded as $\varepsilon \to 0$.

Let us show that $(d^\varepsilon)^T \nabla^2 \Phi_j^\varepsilon(z^\varepsilon) d^\varepsilon$ tends to $-\infty$ as $\varepsilon \to 0$. To simplify the notation, we denote $s^\varepsilon \equiv G_j(z^\varepsilon)$ and $t^\varepsilon \equiv H_j(z^\varepsilon)$. Note that $s^\varepsilon > 0$ and $t^\varepsilon > 0$ by Lemma 3.1, and $s^\varepsilon \to 0$ and $t^\varepsilon \to 0$ since $j \in \bar{\beta}$. Then, by Lemma 3.1 and (17), we can write

$$(d^\varepsilon)^T \nabla^2 \Phi_j^\varepsilon(z^\varepsilon) d^\varepsilon = \frac{\partial \phi_j^\varepsilon}{\partial a}(d^\varepsilon)^T \nabla^2 G_j(z^\varepsilon) d^\varepsilon + \frac{\partial \phi_j^\varepsilon}{\partial b}(d^\varepsilon)^T \nabla^2 H_j(z^\varepsilon) d^\varepsilon - R^\varepsilon, \quad (18)$$

where

$$R^\varepsilon \equiv \frac{t^\varepsilon(t^\varepsilon + 2s^\varepsilon)}{(s^\varepsilon + t^\varepsilon)^3}\left(\frac{s^\varepsilon}{s^\varepsilon + t^\varepsilon}\right)^2 + \frac{2s^\varepsilon t^\varepsilon}{(s^\varepsilon + t^\varepsilon)^3}\left(\frac{s^\varepsilon}{s^\varepsilon + t^\varepsilon}\right)\left(\frac{t^\varepsilon}{s^\varepsilon + t^\varepsilon}\right)$$
$$+ \frac{s^\varepsilon(s^\varepsilon + 2t^\varepsilon)}{(s^\varepsilon + t^\varepsilon)^3}\left(\frac{t^\varepsilon}{s^\varepsilon + t^\varepsilon}\right)^2$$
$$= \frac{2s^\varepsilon t^\varepsilon}{(s^\varepsilon + t^\varepsilon)^3}. \quad (19)$$

By the asymptotic weak nondegeneracy of z^ε, there exist $\rho_1, \rho_2 > 0$ such that $0 < \rho_1 \leq s^\varepsilon / t^\varepsilon \leq \rho_2 < \infty$ for all $\varepsilon > 0$. Therefore, as $\varepsilon \to 0$, we have

$$\frac{2s^\varepsilon t^\varepsilon}{(s^\varepsilon + t^\varepsilon)^3} = \frac{1}{t^\varepsilon} \cdot \frac{2(s^\varepsilon / t^\varepsilon)}{[(s^\varepsilon / t^\varepsilon) + 1]^3} \to \infty$$

and hence, by (18),

$$(d^\varepsilon)^T \nabla^2 \Phi_j^\varepsilon(z^\varepsilon) d^\varepsilon \;\to\; -\infty.$$

Because $(d^\varepsilon)^T \nabla^2 f(z^\varepsilon) d^\varepsilon$, $u_i^\varepsilon (d^\varepsilon)^T \nabla^2 \Phi_i^\varepsilon(z^\varepsilon) d^\varepsilon \, (i \neq j)$, $\lambda_l^\varepsilon (d^\varepsilon)^T \nabla^2 g_l(z^\varepsilon) d^\varepsilon$ and $\mu_k^\varepsilon (d^\varepsilon)^T \nabla^2 h_k(z^\varepsilon) d^\varepsilon$ are bounded as $\varepsilon \to 0$, it follows from (16) that $\bar{u}_j < 0$ implies

$$(d^\varepsilon)^T \nabla_z^2 \mathcal{L}(z^\varepsilon, u^\varepsilon, \lambda^\varepsilon, \mu^\varepsilon) d^\varepsilon \;\to\; -\infty. \tag{20}$$

However, since $(z^\varepsilon, u^\varepsilon, \lambda^\varepsilon, \mu^\varepsilon)$ satisfies, by assumption, the second-order necessary conditions (12), $d^\varepsilon \in \mathcal{T}^\varepsilon(z^\varepsilon)$ implies

$$(d^\varepsilon)^T \nabla_z^2 \mathcal{L}(z^\varepsilon, u^\varepsilon, \lambda^\varepsilon, \mu^\varepsilon) d^\varepsilon \;\geq\; 0.$$

Since this inequality holds for each $\varepsilon > 0$ and d^ε is bounded, we have a contradiction with (20). Hence we have (14).
Consequently, by putting

$$\begin{aligned}
v_i &= \begin{cases} \bar{u}_i, & i \in \bar{\alpha} \\ \bar{u}_i \bar{\xi}_i, & i \in \bar{\beta} \end{cases} \\
w_i &= \begin{cases} \bar{u}_i \bar{\eta}_i, & i \in \bar{\beta} \\ \bar{u}_i, & i \in \bar{\gamma} \end{cases} \\
\lambda_l &= \bar{\lambda}_l, \quad l \in \bar{\mathcal{I}} \\
\mu_k &= \bar{\mu}_k, \quad k \in \mathcal{J},
\end{aligned}$$

we deduce that (13) yields the KKT conditions (2) for the relaxed problem $\mathrm{R}(\bar{z})$. The last part of the theorem follows from Proposition 2.1. □

Acknowledgments We thank Nobuo Yamashita and Junji Imai for their comments on an earlier version of the paper.

References

[1] Y. Chen and M. Florian, "The nonlinear bilevel programming problem: Formulations, regularity and optimality conditions," *Optimization* 32 (1995) 193–209.

[2] F. Facchinei, H. Jiang and L. Qi, "A smoothing method for mathematical programs with equilibrium constraints," *Mathematical Programming*, forthcoming.

[3] A. Fischer, "A special Newton-type optimization method," *Optimization* 24 (1992) 269–284.

[4] M. Fukushima, Z.-Q. Luo and J.-S. Pang, "A globally convergent sequential programming algorithm for mathematical programs with linear complementarity constraints," *Computational Optimization and Applications* 10 (1998) 5–30.

[5] H. Jiang and D. Ralph, "Smooth SQP methods for mathematical programs with nonlinear complementarity constraints," Working paper, Department of Mathematics and Statistics, University of Melbourne, Parkville, Australia, 1997.

[6] C. Kanzow, "Some noninterior continuation methods for linear complementarity problems," *SIAM Journal on Matrix Analysis and Applications* 17 (1996) 851–868.

[7] Z.-Q. Luo, J.-S. Pang and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, (Cambridge University Press, Cambridge, 1996).

[8] J.V. Outrata, "Optimality conditions for a class of mathematical programs with equilibrium constraints," *Mathematics of Operations Research*, forthcoming.

[9] J. Outrata, M. Kočvara and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, (Kluwer Academic Publishers, Dordrecht, 1998).

[10] J. Outrata and J. Zowe, A numerical approach to optimization problems with variational inequality constraints, *Mathematical Programming* 68 (1995) 105–130.

[11] J.-S. Pang and M. Fukushima, "Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints," *Computational Optimization and Applications*, forthcoming.

[12] H. Scheel and S. Scholtes, "Mathematical programs with equilibrium constraints: Stationarity, optimality, and sensitivity," Working paper, The Judge Institute of Management Studies, University of Cambridge, Cambridge, 1997.

[13] J.J. Ye, "Optimality conditions for optimization problems with complementarity constraints," *SIAM Journal on Optimization*, forthcoming.